

2051498-储岱泽-假设检验

[Code ▾](#)

This is an R Markdown (<http://rmarkdown.rstudio.com>) Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Cmd+Shift+Enter*.

导入数据：

[Hide](#)

```
#Body piercing data
american.bp<-c(3,5,2,1,4,4,6,3,5,4)
european.bp<-c(6,5,7,7,6,3,4,6,5,4)
#Store data in a dataframe
bp.survey <- data.frame("bp"=c(american.bp,european.bp),
                        "group"=rep(c("American","European"),each=10),
                        stringsAsFactors = FALSE)
```

[Hide](#)

```
# 加载Hypothesis_pirates.RData文件中的数据
load("Hypothesis_ pirates.RData")
# 查看数据结构
str(bp.survey)
```

```
'data.frame':   20 obs. of  2 variables:
 $ bp    : num  3 5 2 1 4 4 6 3 5 4 ...
 $ group: chr  "American" "American" "American" "American" ...
```

[Hide](#)

```
# 显示数据
bp.survey
```

bp	group
<dbl>	<chr>
3	American
5	American
2	American
1	American
4	American
4	American
6	American
3	American
5	American

bp group

<dbl> <chr>

4 American

1-10 of 20 rows

Previous 1 2 Next

步骤 1:

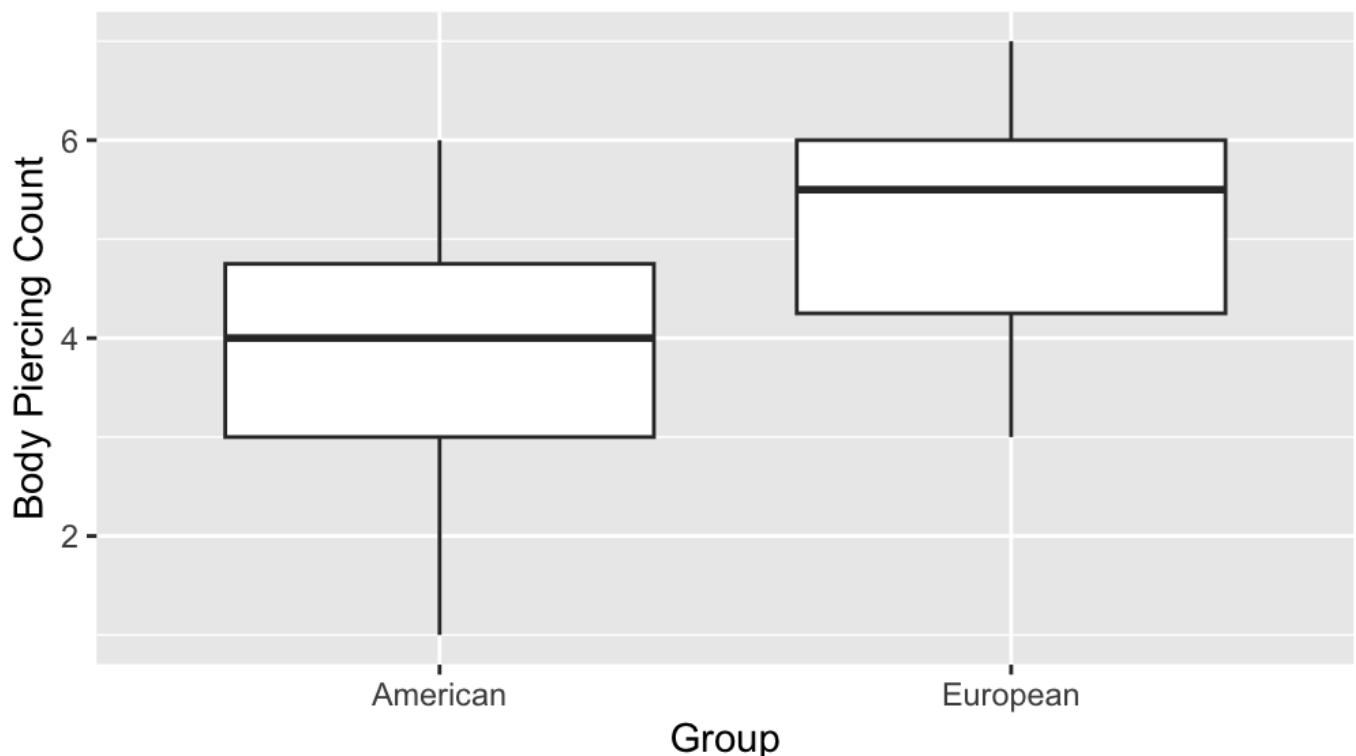
对海盗身体穿孔数目（按照美国，欧洲进行分组）进行箱线图绘图，解读图，阐述美国 和欧洲是否有差异？

Hide

```
# 安装ggplot2包（如果你还没有安装）
#install.packages("ggplot2")

# 加载ggplot2包
library(ggplot2)
# 绘制箱线图
ggplot(bp.survey, aes(x = group, y = bp)) +
  geom_boxplot() +
  xlab("Group") +
  ylab("Body Piercing Count") +
  ggtitle("Body Piercing Count of American and European Pirates")
```

Body Piercing Count of American and European Pirates



对箱线图进行进一步分析，分别计算出箱线图上美国和欧洲的海盗身体打洞数据的最小值、下四分位数、中位数、上四分位数和最大值。

Hide

```
american<-subset(bp.survey,group=='American')
american<-american.bp
y<-boxplot.stats(american,coef=1.5,do.conf=TRUE,do.out=TRUE)
#boxplot(american) #绘制箱线图y$statsy$out

cat("美国海盗身体穿孔数目最小值: ",y$stats[1],'\n')
```

美国海盗身体穿孔数目最小值: 1

Hide

```
cat("美国海盗身体穿孔数目下四分位数: ", y$stats[2],'\n')
```

美国海盗身体穿孔数目下四分位数: 3

Hide

```
cat("美国海盗身体穿孔数目中位数: ", y$stats[3],'\n')
```

美国海盗身体穿孔数目中位数: 4

Hide

```
cat("美国海盗身体穿孔数目上四分位数: ", y$stats[4],'\n')
```

美国海盗身体穿孔数目上四分位数: 5

Hide

```
cat("美国海盗身体穿孔数目最大值: ", y$stats[5],'\n')
```

美国海盗身体穿孔数目最大值: 6

Hide

```
y$out
```

```
numeric(0)
```

Hide

```
european<-subset(bp.survey,group=='European')
european<-european.bp
x<-boxplot.stats(european,coef=1.5,do.conf=TRUE,do.out=TRUE)
cat("欧洲海盗身体穿孔数目最小值: ",x$stats[1],'\n')
```

欧洲海盗身体穿孔数目最小值: 3

Hide

```
cat("欧洲海盗身体穿孔数目下四分位数: ", x$stats[2], '\n')
```

欧洲海盗身体穿孔数目下四分位数: 4

Hide

```
cat("欧洲海盗身体穿孔数目中位数: ", x$stats[3], '\n')
```

欧洲海盗身体穿孔数目中位数: 5.5

Hide

```
cat("欧洲海盗身体穿孔数目上四分位数: ", x$stats[4], '\n')
```

欧洲海盗身体穿孔数目上四分位数: 6

Hide

```
cat("欧洲海盗身体穿孔数目最大值: ", x$stats[5], '\n')
```

欧洲海盗身体穿孔数目最大值: 7

Hide

```
x$out
```

```
numeric(0)
```

通过箱线图以及计算，我们可以发现美国海盗普遍不如欧洲海盗更喜欢穿孔，因为欧洲海盗的穿孔平均值，中位数，最小值以及最大值都比美国海盗更高；同时欧洲海盗的数据分布更加集中。

步骤二：

使用 t 检验（t-test）判断美国和欧洲海盗身上穿孔数目是否有差异，零假设是没有差异，备择假设是有差异，请将 t-test 结果保存在 R 变量 p.test 中，查看 p.test 对象内容，依据对象内容来回答“美国和欧洲海盗身上穿孔数目是否有差异”这个问题。

Hide

```
# 提取美国海盗和欧洲海盗的身体穿孔数据
american <- subset(bp.survey, group == 'American')$bp
european <- subset(bp.survey, group == 'European')$bp

# 执行 t 检验
p.test <- t.test(american, european)

# 查看 p.test 对象内容
print(p.test)
```

Welch Two Sample t-test

```
data: american and european
t = -2.5228, df = 17.783, p-value = 0.0214
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.9335927 -0.2664073
sample estimates:
mean of x mean of y
    3.7      5.3
```

根据 t 检验的结果，p 值为 0.0214，小于显著性水平 0.05。因此，有足够的证据拒绝零假设，即可以认为美国和欧洲海盗身上穿孔数目存在差异。该结果表明，在样本中，美国海盗的平均穿孔数目为 3.7，而欧洲海盗的平均穿孔数目为 5.3。置信区间为 [-2.9335927, -0.2664073]。综上所述，根据 t 检验的结果，可以得出结论：美国和欧洲海盗身上的穿孔数目存在差异。

##步骤三 使用 t 检验 (t-test) 判断 29 岁和 30 岁海盗的文身情况是否有差异，零假设是没有差异，备择假设是有差异，请将 t-test 结果保存在 R 变量 t.test 中，查看 t.test 对象内容，依据对象内容来回答“29 岁和 30 岁海盗的文身情况是否有差异”这个问题。

Hide

```
# 29岁和30岁海盗的纹身情况数据分别保存在 age_29 和 age_30 中
age_29<-subset(pirates,pirates$age==29)$tattoos
age_30<-subset(pirates,pirates$age==30)$tattoos
# 执行 t 检验
t.test.result <- t.test(age_29, age_30)

# 查看 t.test 对象内容
print(t.test.result)
```

Welch Two Sample t-test

```
data: age_29 and age_30
t = 0.26552, df = 119.15, p-value = 0.7911
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.058586  1.386455
sample estimates:
mean of x mean of y
10.081967  9.918033
```

根据 t 检验的结果，p 值为 0.7911，远远大于通常设定的显著性水平 0.05。因此，根据统计检验结果，我们没有足够的证据拒绝零假设，即认为 29 岁和 30 岁海盗的纹身情况在样本中没有显著差异。具体来说，30 岁海盗的平均纹身状况评分为 9.92，而29岁海盗的平均纹身状况为10.08。置信区间为 [-1.058586, 1.386455]。综上所述：根据 t 检验的结果，可以得出结论：29 岁和 30 岁海盗的纹身情况在样本中没有表现出显著差异。

步骤四

卡方检验(Chi-square test)可以用来判断一个因子变量是否依赖另一个因子变量，请考察一个海盗上的大学和海盗是否佩戴眼罩是否有关联，使用卡方检验，将结果保存在 R 变量 c.test 中，查看 c.test 对象内容，依据对象内容来回答“一个海盗上的大学和海盗是否佩戴眼罩是否有关联”这个问题。

```
# 假设这里有相应的数据，海盗上的大学信息保存在 university 中，佩戴眼罩的信息保存在 eyepatch 中
university<-pirates$college
eyepatch<-pirates$eyepatch
#university
#eyepatch
# 执行卡方检验
c.test <- chisq.test(university, eyepatch)

# 查看 c.test 对象内容
print(c.test)
```

```
Pearson's Chi-squared test with Yates' continuity
correction
```

```
data: university and eyepatch
X-squared = 0, df = 1, p-value = 1
```

根据卡方检验的结果，p 值为 1，这意味着我们没有足够的证据拒绝零假设。所以，在这个样本中，海盗上的大学和是否佩戴眼罩之间似乎没有显著的关联。我们无法得出海盗上的大学和是否佩戴眼罩之间存在关联的结论。

Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Cmd+Option+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Cmd+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.